

(11)特許出願公開番号  
特開2003-323580  
(P2003-323580A)

F1			テ-マ-ト <sup>+</sup> (参考)
C06F	19/00	300B	5L096
C06F	7/60	200K	

(71)出願人 000003223  
富士通株式会社  
神奈川県川崎市中原区上小田中4丁目1番  
1号

(72)発明者 橋本 幸治  
神奈川県川崎市中原区上小田中4丁目1番  
1号 富士通株式会社内

(72)発明者 小泉 樹一  
神奈川県川崎市中原区上小田中4丁目1番  
1号 富士通株式会社内

(74)代理人 100072590  
弁理士 井桁 貞一

[illegible]

## 【特許請求の範囲】

【請求項1】 帳票画面用電子フォーム作成装置において、文字記入領域を有する既存の紙帳票を模した帳票画面上的前記文字記入領域に対応するフィールド情報を作成する情報処理をコンピュータに実行させるフィールド情報作成プログラムであって、前記紙帳票上の文字列と罫線に対応する情報を格納した文字列・罫線データベースに基づいて、独立した水平な線分を抽出する独立水平線分抽出段階と、該抽出された独立水平線分に基づいて、フィールド位置座標とフィールド幅とを決定したフィールド候補を作成するフィールド候補生成段階と、をコンピュータに実行させることを特徴とするフィールド情報作成プログラム。

【請求項2】 さらに、前記フィールド候補のフィールド位置座標とフィールド幅で決まる領域に複数文字列が所定値以上の間隔で存在することを判定する文字列間隔判定段階と、該判定された間隔に対応したフィールド位置座標とフィールド幅とを決定したフィールド候補を生成するフィールド候補生成段階と、をコンピュータに実行させることを特徴とする請求項1記載のフィールド情報作成プログラム。

【請求項3】 さらに、前記独立水平線分の上部の空白領域の高さを判定する上部空白判定段階と、該判定された空白の高さに基づいて該フィールド候補のフィールド高さを設定するフィールド高さ設定段階と、をコンピュータに実行させることを特徴とする請求項1または請求項2記載のフィールド情報作成プログラム。

【請求項4】 帳票画面用電子フォーム作成装置において、文字記入領域を有する既存の紙帳票を模した帳票画面上的前記文字記入領域に対応するフィールド情報を作成するフィールド情報作成方法であって、前記紙帳票上の文字列と罫線に対応する情報を格納した文字列・罫線データベースに基づいて、独立した水平な線分を抽出する独立水平線分抽出段階と、該抽出された独立水平線分に基づいて、フィールド位置座標とフィールド幅とを決定したフィールド候補を作成するフィールド候補生成段階と、を有することを特徴とするフィールド情報作成方法。

【請求項5】 帳票に関する情報処理をコンピュータに実行させるための帳票定義データである帳票画面用電子フォームを作成する帳票画面用電子フォーム作成装置であって、未記入紙帳票の文字列と罫線に対応する情報を格納した文字列・罫線データベースに基づいて、独立した水平な線分を抽出する独立水平線分抽出手段と、該抽出された独立水平線分に基づいて、フィールド位置座標とフィールド幅とを決定したフィールド候補を作成するフィールド候補生成手段と、を有することを特徴と

する帳票画面用電子フォーム作成装置。

## 【発明の詳細な説明】

## 【0001】

【発明の属する技術分野】本発明は、帳票に関する情報処理をコンピュータに実行させるための帳票定義データである帳票画面用電子フォームを作成する帳票画面用電子フォーム作成技術に係り、特に、文字記入領域を有する既存の紙帳票を模した帳票画面上的前記文字記入領域に対応するフィールド情報を作成するフィールド情報作成技術に関する。

## 【0002】

【従来の技術】IT技術の進展によって、企業内、企業間にとどまらず、行政、消費者を含む多くの分野で、従来の紙帳票に代わり、帳票画面を端末から扱う電子帳票システムがますます必要となってきた。このようなシステムの設計者は、帳票に関する情報処理をコンピュータに実行させるための帳票定義データである帳票画面用電子フォームを設計するが、このとき、設計の効率化の観点からも、また、エンドユーザの立場からも、できるだけ既存の紙帳票と等しいデザインの帳票画面が望ましい。

【0003】このような帳票画面用電子フォーム作成において、未記入紙帳票のビットマップイメージデータを利用して、文字記入領域を有する既存の紙帳票を模した帳票画面上的前記文字記入領域に対応するフィールド情報を自動生成する従来技術がある。例えば、特許第3028833号特許公報においては、少なくとも3辺が線パターンによって囲まれた矩形領域やプレプリント文字列を前記ビットマップイメージデータから抽出することにより、文字記入領域に対応するフィールド情報を自動生成する技術が開示されている。

【0004】本公知例に基づいて、従来の帳票画面用電子フォーム作成装置とフィールド情報について、図9によって説明する。図9(a)は従来の帳票画面用電子フォーム作成装置の構成例である。文字列・罫線抽出手段1は未記入紙帳票イメージデータファイルimage から文字列および罫線を認識して、文字列の値、位置、フォントサイズ、罫線の太さ、種類、位置、長さなどを文字列・罫線データベースMKDBに格納する。従来のフィールド候補生成手段2'はこの文字列・罫線データベースMKDBをもとにフィールド情報mfを抽出する。

【0005】フィールド情報mfとは、文字記入領域を有する既存の紙帳票を模した帳票画面上的前記文字記入領域に対応する情報であって、そのデータ構造例を図9(b)に示すように、フィールドIDfid、フィールド左下座標(x,y)、フィールド幅w、フィールド高さhなどからなる。従来のフィールド候補生成手段2'では既存の紙帳票で記入領域あらずキーワードをキーワードデータベースkdDBとしてもち、記入領域はこのキーワードの近辺にある矩形または空白であるとする。例えばフ

フィールド候補生成手段2'の矩形領域抽出手段22'は図8(a)のイメージを持つ未記入紙帳票イメージデータファイルimageに基づき文字列・罫線データベースMDBから矩形領域b81を抽出する。またフィールド候補生成手段2'のキーワード照合手段24'は文字列・罫線データベースMDBから「年」「月」「日」「住所」「都道府県」「市郡」「区」をそれぞれキーワードデータベースMDBに合致するキーワードとして抽出する。文字列間隔判定手段23'は矩形領域抽出手段22'の抽出した矩形内の文字列間の間隔を判定し、図8(a)の場合は4つの文字列間隔に対応するフィールド情報mfを生成する。これは図8(b)の帳票画面におけるb85~b88に対応するものである。この結果、矩形領域抽出手段22'は図8(b)の文字入力領域b81に対応するフィールド情報mf(フィールドIDfid2')を、文字列間隔判定手段23'は文字入力領域b85~b88に対応する4つのフィールド情報mf(フィールドIDfid3')を、またキーワード照合手段24'は文字入力領域b82~b84に対応する3つのフィールド情報mf(フィールドIDfid3')を生成する。ここにこれらフィールド情報mfは各文字入力領域に対応するフィールド左下端座標(x,y)、フィールド幅w、フィールド高さhがフィールド候補生成手段2'によって決定されている。そして重なり整理手段3'はこれらのフィールド情報mfのフィールド左下端座標(x,y)、フィールド幅wからこれらのフィールド情報mfのうち重なっている文字入力領域b81に対応のフィールド情報mfを削除して、これらフィールド情報mfを第1のフィールド情報データベースMDB1'に記入する。

【0006】紙帳票にユーザが記入する他の領域には選択肢領域や付添付領域などがある。例えば、図8(a)の紙帳票の「都道府県」は該当に○を付する選択肢である。図8(b)の帳票画面の例ではこれはスクロール選択肢オブジェクトsoとして実現されている。電子フォーム生成手段6'では第1のフィールド情報データベースFDB1'のフィールド情報mfの修正や、上記のような選択肢オブジェクトsoの設定、写真やサインのようなイメージ領域オブジェクトicの設定を設計者と対話的に行う。このとき、未記入紙帳票イメージデータファイルimageからは紙帳票にプレプリントされたロゴや飾りのイメージを取り込み、また文字列・罫線データベースMDBからプレプリント文字列・罫線を取り込み、最終的な帳票画面用電子フォームE-formを完成する。

【0007】

【発明が解決しようとする課題】上記の従来の帳票画面用電子フォーム作成装置では、矩形領域とキーワードに基づいてフィールド情報mfを設定していた。しかし、キーワードを事前登録していない紙帳票や矩形によらない記入指示の紙帳票では、従来の帳票画面用電子フォーム作成装置ではフィールド情報mfを的確に生成することができない、という課題があった。また、文字列・罫線抽

出手段1が文字列抽出に失敗してキーワードが得られなかった場合にも、矩形以外の字がかりがないと設定すべきフィールド情報mfが生成できない。例えば、図7のような下線を多用した紙帳票の場合には、図のA部以外で、キーワードデータベースMDBにキーワード登録されていないところはフィールド情報mfが設定されない。また、B部のように、仮にWhichがキーワード登録されていたとしても、b1~b9をこのような位置でフィールド情報mfとして生成することはできない。本発明はこのような課題に鑑み、下線によって文字記入領域を指示するような紙帳票を模した帳票画面用電子フォーム作成装置を提供することを目的とする。

【0008】

【課題を解決するための手段】上記課題は、図2の動作フロー図に示す如く、帳票画面用電子フォーム作成装置において、文字記入領域を有する既存の紙帳票を模した帳票画面用の前記文字記入領域に対応するフィールド情報を作成するフィールド情報作成方法であって、前記紙帳票上の文字列と罫線に対応する情報を格納した文字列・罫線データベースMDBに基づいて、独立した水平な線分を抽出する独立水平線分抽出段階(図2のステップS202)と、該抽出された独立水平線分に基づいて、フィールド左下端座標(x1,y1)とフィールド幅w1とを決定したフィールド候補(フィールドIDfid1)を作成するフィールド候補生成段階(図2のステップS202)と、を有するフィールド情報作成方法によって解決される。

【0009】すなわち、例えば図7のような未記入紙帳票イメージデータファイルimageから得られた文字列・罫線データベースMDBでは、両端近傍が空白で両端y座標が等しい罫線が独立水平線分として抽出され、これらの多くは文字記入領域であるので、たとえキーワード設定が不十分であっても、従来の帳票画面用電子フォーム作成装置では得られなかった図7のB部に対応するようなフィールド情報mfが作成可能となる。

【0010】さらに、上記課題は、前記フィールド候補(フィールドIDfid1)のフィールド左下端座標(x1,y1)とフィールド幅w1で決まる領域に複数文字列が所定値以上の間隔で存在することを判定する文字列間隔判定段階(図2のステップS203)と、該判定された間隔に対応したフィールド左下端座標(x3,y3=y2)とフィールド幅w3とを決定したフィールド候補(フィールドIDfid3)を生成するフィールド候補生成段階(図2のステップS203)と、を有するフィールド情報作成方法によって解決される。

【0011】すなわち、文字列間隔判定段階では、図7のC部のように下線上に文字列Date,Offense,Court,Dispositionのように複数文字列が間隔を空けて配置されていることを検出するので、c1~c4に対応するフィールド

情報mfを的確に作成することができる。さらに、上記課題は、前記独立水平線分(第1のフィールド情報データベースFD01の各フィールドI D fid)の上部の空白領域の高さを判定する上部空白判定段階(図2のステップS210)と、該判定された空白の高さに基づいて該フィールド候補のフィールド高さhを設定するフィールド高さ設定段階(図2のステップS210)と、を有するフィールド情報作成方法によって解決される。

【0012】すなわち、下線によるフィールド情報mf作成では矩形領域による場合と異なりフィールド高さhを決める基準が必要であり、一般に上部空白によってその高さを設定すれば、プレプリント部分と重なることのないフィールド情報mfを的確に作成することができる。さらに、上記課題は、前記フィールド候補(第1のフィールド情報データベースFD01の各フィールドI D fid)のフィールド左下端座標(x,y)とフィールド幅wで決まる領域の左右の所定範囲に存在する文字列の高さを判定する左右部文字列判定段階(図2のステップS210)と、該判定された文字列の高さに基づいて該フィールド候補のフィールド高さhを設定するフィールド高さ設定段階(図2のステップS210)と、を有するフィールド情報作成方法によって解決される。すなわち、文字記入領域の状況によっては上部空白が高くても、左右プレプリント文字フォントの高さに合わせた文字記入領域を予定する場合もあるので、このような場合、左右部文字列判定段階によってフィールド高さhを正しく設定したフィールド情報mfを的確に作成することができる。

【0013】さらに、上記課題は、前記フィールド候補(第2のフィールド情報データベースFD02の各フィールドI D fid)のフィールド下端座標yが等しい複数のフィールド候補のフィールド高さhを比較して、これら複数のフィールド候補のフィールド高さhを同一値に揃えるフィールド高さ調整段階(図2のステップS211)を有するフィールド情報作成方法によって解決される。

【0014】すなわち、図7のB部におけるb1~b3の部分は上部空白が大きく、b1~b3の部分は上部空白が小さいが、このB部の文字記入領域はすべて同一高さであることが期待されている、と考えられる。このような場合、フィールド高さ調整段階においておのおののフィールド高さhの分布に基づいて、的確にこれら複数のフィールド候補のフィールド高さhを同一値に揃えることができる。

【0015】さらに、上記課題は、前記生成されたフィールド候補(図2のステップS206で得られたフィールドI D fid3およびステップS208で得られたフィールドI D fid4)のフィールド左下端座標(x,y)およびフィールド幅wに基づいて、重なったフィールド候補を判定して削除する重なり整理段階(図2のステップS209)を有するフィールド情報作成方法によって解決される。

【0016】すなわち、例えば図1のキーワードデータベースKDBに文字列dateが登録されている場合、図7のC部における文字記入領域clは、ステップS205によってステップS208によってのいずれによってもフィールド情報mfを生成する。このような場合、重なり整理段階で検出され、所定の選択はジョックで一つの的確に選択される。

【0017】

【発明の実施の形態】本発明のフィールド情報作成方法の実施例を図1~図9により説明する。なお、本発明におけるコンピュータ処理は、当該コンピュータの主記憶装置上に展開されたコンピュータプログラムにより実行されるが、このコンピュータプログラムの提供形態は、当該コンピュータに接続された補助記憶装置をはじめ、CD-ROM等の可搬型記憶装置やネットワーク接続された他のコンピュータの主記憶装置及び補助記憶装置等の各記録媒体に格納されて提供されるもので、このコンピュータプログラムの実行に際しては、当該コンピュータの主記憶装置上にローディングされ実行されるものである。

【0018】図1は本発明の帳票画面用電子フォーム作成装置の一例の実施例の構成を示すものである。この例では、図9の従来の帳票画面用電子フォーム作成装置と対比できるように、同一機能の要素は同一番号を付して示した。図1と図9を対比してわかるように、本発明では、従来のフィールド候補生成手段2'、重なり整理手段3'に相当するフィールド候補生成手段2、重なり整理手段3に加えて、フィールド高さ設定手段4およびフィールド高さ調整手段5が付加されている。また、フィールド候補生成手段2を従来のフィールド候補生成手段2'と対比すると、独立水平線分抽出手段21が新規に付加され、矩形領域抽出手段22、文字列間隔判定手段23、キーワード照合手段24は従来と同等(であるが少し異なる)機能である。すなわち、図9の従来のフィールド候補生成手段2'ではすでに説明した如く、フィールド情報mfとしてフィールドI D fid'、フィールド左下端座標(x,y)、フィールド幅wに加えてフィールド高さhをも決定したが、フィールド候補生成手段2においてはフィールド高さhは決定されない。そして、フィールド高さhはフィールド高さ設定手段4およびフィールド高さ調整手段5によって決定される。

【0019】図1において図9で説明していない部分を説明すると、フィールド候補生成手段2の独立水平線分抽出手段21は文字列・罫線データベースKDBから独立水平線分を抽出して、そのフィールドI D fid1、フィールド左下端座標(x1,y1)、フィールド幅w1を文字列間隔判定手段23に与える。文字列間隔判定手段23は独立水平線分抽出手段21および矩形領域抽出手段22からのフィールド情報mfを受け、文字列間隔判定を行い、新たなフィールドI D fid3、フィールド左下端座標(x3,y3)フィールド

ド幅w3からなるフィールド情報mfを重なり整理手段3に与える。重なり整理手段3はキーワード照合手段24からのフィールドIDfid4、フィールド左下端座標(x4,y4)、フィールド幅w4からなるフィールド情報mfを変えて重なり整理を行い、その結果を第1のフィールド情報データベースFDB1に格納する。この時点で第1のフィールド情報データベースFDB1は従来の第1のフィールド情報データベースFDB1'と異なり、フィールド高さhの情報を欠いたフィールド情報mfが格納されている。

【0020】フィールド高さ設定手段4は上部空白判定手段41および左右部文字列判定手段42からなる。これらの手段とともに第1のフィールド情報データベースFDB1の各フィールドIDfid1に対してフィールド高さhの設定を行い、第2のフィールド情報データベースFDB2に格納する。フィールド高さ調整手段5は第2のフィールド情報データベースFDB2に格納された各フィールドIDfid1に対してフィールド高さhの調整を行い、最終的なフィールド情報mfとして第3のフィールド情報データベースFDB3にこれを格納する。これによって従来の第1のフィールド情報データベースFDB1'と等価な第3のフィールド情報データベースFDB3が得られ、従来と同様に電子フォーム生成手段6がこれらをもとに帳票画面用電子フォームE-formを生成する。

【0021】次に、図2の動作フローにより、本発明の帳票画面用電子フォーム作成装置の動作を説明する。まずステップS201において、文字列・罫線抽出手段1が未記入紙帳票イメージデータファイルimageからものと紙帳票のフレプリント文字列・罫線を抽出し文字列・罫線データベースMKDBに格納する。これは従来の帳票画面用電子フォーム作成装置と同様である。図3には文字列・罫線データベースMKDBのデータ構造例を示す。図3(a)は文字列の例で、各文字列の位置、左右方向の幅、そのフォントサイズに対応した高さ(hで表示)を格納する。図3(b)は罫線の例で、罫線の位置、左右方向の幅が与られている。これら文字列・罫線の抽出方法については、前記の特許第3028833号特許公報に示される技術や類似の方法が公知である。

【0022】図2に展って、次にステップS202において、フィールド候補生成手段2の独立水平線分抽出手段21は、文字列・罫線データベースMKDBから独立水平線分を抽出してフィールドIDfid1、フィールド左下端座標(x1,y1)、フィールド幅w1を設定する。独立水平線分とは、その両端に他の罫線との交点をもたず、かつ、その両端のy座標が等しい(スローフをもたない)罫線であり、その抽出方法の詳細については後に図4によって詳述する。

【0023】抽出された線分に対応するフィールド情報mf候補として独立水平線分抽出手段21はフィールドIDfid1を発行し、この線分によって定まるフィールド左下端座標(x1,y1)、フィールド幅w1をバッファ上に設定し

て文字列間隔判定手段23を起動する。ここにフィールド左下端座標(x1,y1)、フィールド幅w1は罫線線分の値そのものではなく、帳票画面上の文字送りピッチ(x方向px:y方向py)で正規化された値をとってもよく、その方法についても前記特許第3028833号特許公報と同様な技術による。

【0024】次に図2のステップS203において、文字列間隔判定手段23は前記バッファ上のフィールドIDfid1に対して、所定探索高さ(例えば対象罫線の上部にマッチ分pyの2倍)の範囲の文字列についてその間隔を判定しフィールドIDfid3、フィールド左下端座標(x3,y3)、フィールド幅w3を設定する。すなわち、文字列間隔判定手段23は図3のような文字列・罫線データベースMKDB中の各文字列の位置座標と文字列幅、高さで決まる文字列領域が領域(x1,y1)-(x1+w1,y1+py)と重なる部分をもつかを检查する。このような検査で複数の文字列が得られた場合は、この罫線内に複数の文字列によって分割される複数のフィールドを設定する。この場合、文字列間隔値の大きさによってフィールドを設定する基準や当該文字列の左、右のいずれをフィールドとするか、という知識が必要であるが、これらの技術も前記特許第3028833号特許公報に示す方法を用いることができる。このようにして、文字列間隔判定手段23は従来の文字列間隔判定手段23'と同様に複数のフィールドIDfid3をもつフィールド情報mfをバッファ上においてステップS204に移行する。なお、この場合、一般にy3=y1である。

【0025】ステップS204でフィールド候補生成手段2はバッファ上のすべてのフィールドIDfid1に対して前記ステップS203がなされたかを監視して、残りがあれば次のフィールドIDfid1を文字列間隔判定手段23に与えてステップS203にループさせる。そして全フィールドIDfid1に対して終了すると、フィールド候補生成手段2はステップS205において、従来の矩形領域抽出手段22'と類似の矩形領域抽出手段22を起動して、文字列・罫線データベースMKDBから矩形領域を抽出してフィールドIDfid2、フィールド左下端座標(x2,y2)、フィールド幅w2をバッファ上に設定して、前記の文字列間隔判定手段23を起動する。

【0026】文字列間隔判定手段23はステップS206において、フィールドIDfid2に対して文字列間隔を判定し、フィールドIDfid3、フィールド左下端座標(x3,y3)、フィールド幅w3をバッファ上に設定する。このステップおよびステップS207は対象をフィールドIDfid2に換えたのみで前記ステップS203、S204と同様動作を行う。なお、この場合、一般にy3=y2である。

【0027】この結果、ステップS207で全フィールドIDfid2に対して終了となった場合、バッファ上にはフィールドIDfid3の複数のフィールド情報mfが存在する。この状態でフィールド候補生成手段2はステップS

208に移行し、キーワード照合手段24を起動し、文字列・罫線データベースYKDBからキーワードデータベースKwDB登録のキーワードを照合しフィールドIDfid4、フィールド左下端座標(x4,y4)、フィールド幅w4を設定する。この詳細についても従来のキーワード照合手段21と同様であり、前記特許第302833号特許公報に示す方法を用いることができる。この結果、バッファ上にはフィールドIDfid3、およびフィールドIDfid4の複数フィールド情報mfが存在する。

【0028】次いで、重なり整理手段3はステップS209において、バッファ上のフィールドIDfid3、フィールドIDfid4のフィールド左下端座標(x3,y3)、(x4,y4)を調べ、これらがある許容値の範囲で等しい場合、いずれかを優先して他を捨てて結果を第1のフィールド情報データベースFDB1に格納する。この詳細についても従来の重なり整理手段3と同様であり、前記特許第302833号特許公報に示す方法を用いることができる。

【0029】以上によって得られた第1のフィールド情報データベースFDB1をもとに、フィールド高さ設定手段4が、次にステップS210において、まず、上部空白判定手段41を起動して、第1のフィールド情報データベースFDB1の各フィールドIDfidに対し、そのフィールド下端(x,y)-(x+w,y)の上部空白を判定してフィールド高さhを設定する。その詳細については図5によって後述する。次に、フィールド高さ設定手段4は左右部文字列判定手段42を起動して、当該フィールドIDfidの左右に文字列があればその高さを加味してフィールド高さhを設定する。その詳細については図6によって後述する。このようにしてフィールド高さ設定手段4は上部空白判定手段41または左右部文字列判定手段42によって各フィールドIDfidに対するフィールド高さhを設定する。なお、このとき、上部空白判定手段41と左右部文字列判定手段42の両者が異なるフィールド高さhを与える場合は、本実施例ではより小さい値を与えるフィールド高さhを採用する。

【0030】以上によって、ステップS210でフィールド高さhを決定したフィールド情報mfが作られ、フィールド高さ設定手段4はこれを第2のフィールド情報デ

評価手段1-1  $lyc - lys \leq lydif$

評価手段1-2  $abs(aTan((lyc-lys+1)/(lx-lys+1))) < l.l$

のどちらかを用いることが可能である。ここで、lydifおよびl.lは、予め指定可能なパラメータであり、lydifは許容するY座標差、l.lは許容する水平に対する傾きを表す。また、abs(x)はxの絶対値を表し、aTan(x) (xのアークタンジェント)を表す。以上を第1の条件とする。

【0031】次に、当該罫線が矩形などの図形を成さない独立した線分であるかを評価する(評価手段2)。この評価手段として、以下の2つの手段が代表的である。  
評価手段2-1 図4(b)の斜線部は、罫線左端の調査範

データベースFDB2に格納する。次いでフィールド高さ調整手段5はステップS211において、この第2のフィールド情報データベースFDB2中の各フィールドIDfidについて、フィールド左下端Y座標yでソートし、同一Yの複数フィールドIDfidについてフィールド高さhを調整する。これは、図7のB部のように上部空白の高さが異なる同一Y座標のフィールド情報IDfidの場合のように、本来期待されるフィールド高さhは同一Y座標フィールドIDfid間で同一であろう、という仮定を実現する。従って、フィールド高さ調整手段5は最も小さいフィールド高さhを与える値を採用して、これを該当全フィールドIDfidに適用する。そしてその修正結果を第3のフィールド情報データベースFDB3に格納する。

【0031】以上によって従来の帳票画面用電子フォーム作成装置の第1のフィールド情報データベースFDB1と等価な第3のフィールド情報データベースFDB3が形成された。本第3のフィールド情報データベースFDB3には、従来の第1のフィールド情報データベースFDB1には収容されなかった独立水平線分に由来するフィールド情報mfが取り込まれている。最後のステップS212において、電子フォーム生成手段6は、従来の帳票画面用電子フォーム作成装置と同様に、未記入紙帳票イメージデータファイルimage、文字列・罫線データベースYKDB、第3のフィールド情報データベースFDB3に基づいて、対話的にフィールド情報mfの修正、罫線オブジェクトsoの設定、イメージ領域オブジェクトloの設定を行い、帳票画面用電子フォームE-formを完成する。

【0032】次に、図4によって独立水平線分抽出手段21の動作を説明する。左端座標(lxs,lys)、右端座標(lxe,lye)をもつ罫線を幅と長さを持つ矩形領域と考え、画像データから得た罫線データの情報は以下の形式で表すことができる。

$(Lxs, Lys) - (Lxe, Lye)$

図4(a)は、罫線情報の例である。

【0033】まず、罫線の方向が水平であるかどうかを評価する(評価手段1)。評価手段は様々であるが以下の2つが代表的である。

罫線の例である。罫線の左右端点の周囲を調査し、黒画素の量が一定以下なら独立線分と判定する。例えば、msnおよびlhrsをそれぞれ、予め設定できるパラメータとした場合、矩形領域 $(lxs - msn, lys - msn) - (lxs + msn, lye + msn)$ から、矩形領域 $(Lxs, Lys) - (Lxe + msn, Lye)$ を除いた領域にある黒画素数をbleflとすれば、 $blefl \leq lhrs$

を満たす場合に、この端点は他の罫線と接続していないと判定する。

【0035】同様に、矩形領域 $(lxe - msn, lys - msn)$

- (Lxc + mgn, Lys + mgn)から、矩形領域(Lxc - mgn, Lys) - (Lxc, Lys)を除いた領域にある黒画素数をBrightとすれば、

Bright  $\leq$  thre

を満たす場合に、この端点は他の罫線と接続していないと判定する。以上により両端点とも、接続されていないと判定されれば独立線分と判定する。

【0036】評価手段2-2 他の罫線と交差しないものを独立線分と判定する。当該罫線が、他の罫線と交差しなければ独立線分と云えることから、図4(c)に示すように線分A-Bと線分C-Dがある場合、ベクトルOAとベクトルOBの外積をP\_AB、ベクトルOCとベクトルODの外積をP\_CD、ベクトルOBとベクトルODの外積をP\_ODとすれば、P\_ABとP\_ODが同じ符号で且つ $abs(P_{AB}) < abs(P_{OD} - P_{CD})$ 成り立つ場合、線分C-Dと線分A-Bは交差するという既知の事実を利用して同様の評価計算を行い判定を行う。

【0037】以上のいずれかの評価を行うことを第2の条件とする。水平線として第1条件を満たし、且つ独立線分として第2条件を見たした罫線を、下線フィールド候補として認定する。次に図5によって、上部空白判定手段41の動作例を説明する。図5(a)に示す如く、当該下線(Lxs, Lys) - (Lxc, Lys)と、予め設定可能なパラメータであり上方探索距離を表すShによって求められる探索範囲(Lxs, Lys - Sh) - (Lxc, Lys)について、文字以外の図形があるかを調査する。文字以外の図形が存在する場合は、その中で最も下端の位置を検出し、その位置をL1とする。図5(a)の例では文字以外は検出しない、次に当該下線と、別途予め設定可能なパラメータであり、上方探索距離を表すSh2によって求められる探索範囲(Lxs, Lys - Sh2) - (Lxc, Lys)について、文字列があるかを調査する。文字列が存在する場合は、その中で最も下端の位置を検出し、その位置をL2とする。

【0038】次に、フィールドの上端位置Fysを以下の様に設定する。

$$Fys = \min(Lys - Fh, \min(L1, L2))$$

ここで、Fhは罫線に対してフィールド領域がどの程度内側かを示す予め設定可能なパラメータであり、 $\min(A, B)$ はAとBの小さい方の値をとるものとする。これにより、上方の文字列等の図形と重ならず適切なサイズのフィールド生成を実現できる。設定した高さはデフォルトのフィールド高さとして、上方に重なるべき図形が存在しない場合に用いられることになる。図5(b)は、設定したフィールド高さhの例である。

【0039】なお、上方の図形下端位置を検出する際に非文字の図形と文字を付けた理由を説明すると、文字列などは下線の上方直近に自身の項目名として存在する場合があります。この場合に誤って図5(c)の様に自身の項目名文字列を避けて設定してしまうことがないように、別途 $Sh2 < Sh$ を満たすSh2により調査範囲を設け、項目名

文字列は上限値計算の算出に影響を与えないようにするためである。

【0040】次に図6によって左右部文字列判定手段42の動作を説明する。図6(a)に示す如く、予め設定可能なパラメータとして左探索距離を表すSleftおよび右探索距離を表すSrightがあるとする。また、当該下線を(Lxs, Lys) - (Lxc, Lys)とする。左右部文字列判定手段42はまず、左方形範囲(Lxs - Sleft, Lys - Sh3) - (Lxs, Lys)を調査し、当該範囲に一部でも重なる文字を全て検出する。ここで文字を検出した場合は、それらの文字の最上端位置をTitleTopYとする。図6(a)は左方の調査範囲および検出した文字を示す。

【0041】左方で文字列が検出された場合、その文字列の最上端位置TitleTopYが求められていて、予め設定したパラメータとして、余裕距離値を表すYmgnとすれば、フィールドのデフォルト高さは、

$$Ph = Lys - (TitleTopY - Ymgn)$$

として、動的に更新する。左方で文字列が検出できなかった場合は、デフォルト高さは予め設定した値を使うものとする。次に、左方と同様に右方を調査した結果を、左方と併用する。図6(b)は、左方にある文字列を考慮して設定したフィールドの例である。

【0042】(付記1) 帳票画面用電子フォーム作成装置において、文字記入領域を有する既存の紙帳票を模した帳票画面上の前記文字記入領域に対応するフィールド情報を作成する情報処理をコンピュータに実行させるフィールド情報作成プログラムであって、前記紙帳票上の文字列と罫線に対応する情報を格納した文字列・罫線データベースに基づいて、独立した水平な線分を抽出する独立水平線分抽出段階と、該抽出された独立水平線分に基づいて、フィールド位置座標とフィールド幅とを決定したフィールド候補を作成するフィールド候補生成段階と、をコンピュータに実行させることを特徴とするフィールド情報作成プログラム。

【0043】(付記2) さらに、前記フィールド候補のフィールド位置座標とフィールド幅で決まる領域に複数文字列が所定値以上の間隔で存在することを判定する文字列間隔判定段階と、該判定された間隔に対応したフィールド位置座標とフィールド幅とを決定したフィールド候補を生成するフィールド候補生成段階と、をコンピュータに実行させることを特徴とする付記1記載のフィールド情報作成プログラム。

【0044】(付記3) さらに、前記独立水平線分の上部の空白領域の高さを判定する上部空白判定段階と、該判定された空白の高さに基づいて該フィールド候補のフィールド高さを設定するフィールド高さ設定段階と、をコンピュータに実行させることを特徴とする付記1または付記2記載のフィールド情報作成プログラム。

【0045】(付記4) さらに、前記フィールド候補のフィールド位置座標とフィールド幅で決まる領域の左



右の所定範囲に存在する文字列の高さを判定する左右部文字列判定段階と、該判定された文字列の高さに基づいて該フィールド候補のフィールド高さを設定するフィールド高さ設定段階と、をコンピュータに実行させることを特徴とする付記1または付記2記載のフィールド情報作成プログラム。

【0046】（付記5）さらに、前記フィールド候補のフィールド下端座標が等しい複数のフィールド候補のフィールド高さを比較して、これら複数のフィールド候補のフィールド高さを同一値に揃えるフィールド高さ調整段階、をコンピュータに実行させることを特徴とする付記3または付記4記載のフィールド情報作成プログラム。

【0047】（付記6）さらに、前記生成されたフィールド候補のフィールド位置座標およびフィールド幅に基づいて、重なったフィールド候補を判定して削除する重なり整理段階、をコンピュータに実行させることを特徴とする付記1または付記2記載のフィールド情報作成プログラム。

【0048】（付記7）帳票画面用電子フォーム作成装置において、文字記入領域を有する既存の紙帳票を模した帳票画面上の前記文字記入領域に対応するフィールド情報を作成するフィールド情報作成方法であって、前記紙帳票上の文字列と罫線に対応する情報を格納した文字列・罫線データベースに基づいて、独立した水平な線分を抽出する独立水平線分抽出段階と、該抽出された独立水平線分に基づいて、フィールド位置座標とフィールド幅とを決定したフィールド候補を作成するフィールド候補生成段階と、を有することを特徴とするフィールド情報作成方法。

【0049】（付記8）帳票に関する情報処理をコンピュータに実行させるための帳票定義データである帳票画面用電子フォームを作成する帳票画面用電子フォーム作成装置であって、未記入紙帳票の文字列と罫線に対応する情報を格納した文字列・罫線データベースに基づいて、独立した水平な線分を抽出する独立水平線分抽出手段と、該抽出された独立水平線分に基づいて、フィールド位置座標とフィールド幅とを決定したフィールド候補を作成するフィールド候補生成手段と、を有することを特徴とする帳票画面用電子フォーム作成装置。

【0050】

【発明の効果】以上の説明から明らかなように、本発明によれば、紙帳票上の文字列と罫線に対応する情報を格納した文字列・罫線データベースに基づいて、独立した水平な線分を抽出する独立水平線分抽出段階と、該抽出された独立水平線分に基づいて、フィールド位置座標とフィールド幅とを決定したフィールド候補を作成するフィールド候補生成段階とを有するフィールド情報作成方法によって、紙帳票上の下線で示された文字記入領域に対応するフィールド情報を自動生成できる、という効果

がある。また、複数文字列が所定値以上の間隔で存在することを判定する文字列間隔判定段階や、左右の所定範囲に存在する文字列の高さを判定する左右部文字列判定段階や、独立水平線分の上部の空白領域の高さを判定する上部空白判定段階や、フィールド候補のフィールド下端座標が等しい複数のフィールド候補のフィールド高さを比較して、これら複数のフィールド候補のフィールド高さを同一値に揃えるフィールド高さ調整段階や、フィールド候補のフィールド位置座標およびフィールド幅に基づいて、重なったフィールド候補を判定して削除する重なり整理段階、をさらに有することにより、より適切なフィールド情報を自動生成できる、という効果がある。

【図面の簡単な説明】

【図1】本発明の帳票画面用電子フォーム作成装置の実施例構成図

【図2】本発明の帳票画面用電子フォーム作成装置の動作フロー図

【図3】文字列・罫線データベースMKDBのデータ構造例

【図4】独立水平線分抽出手段21の動作例

【図5】上部空白判定手段41の動作例

【図6】左右部文字列判定手段42の動作例

【図7】紙帳票例

【図8】紙帳票と帳票画面の対比例

【図9】従来の帳票画面用電子フォーム作成装置構成例とフィールド情報mfのデータ構造例

【符号の説明】

- 1 文字列・罫線抽出手段
- 2 フィールド候補生成手段
- 3 重なり整理手段
- 4 フィールド高さ設定手段
- 5 フィールド高さ調整手段
- 6 電子フォーム生成手段
- 21 独立水平線分抽出手段
- 22 矩形領域抽出手段
- 23 文字列間隔判定手段
- 24 キーワード照合手段
- 41 上部空白判定手段
- 42 左右部文字列判定手段

image 未記入紙帳票イメージデータファイル

MKDB 文字列・罫線データベース

FDB1 第1のフィールド情報データベース

FDB2 第2のフィールド情報データベース

FDB3 第3のフィールド情報データベース

E-form 帳票画面用電子フォーム

fid フィールドID

(x,y) フィールド左下端座標

w フィールド幅

h フィールド高さ

mf フィールド情報

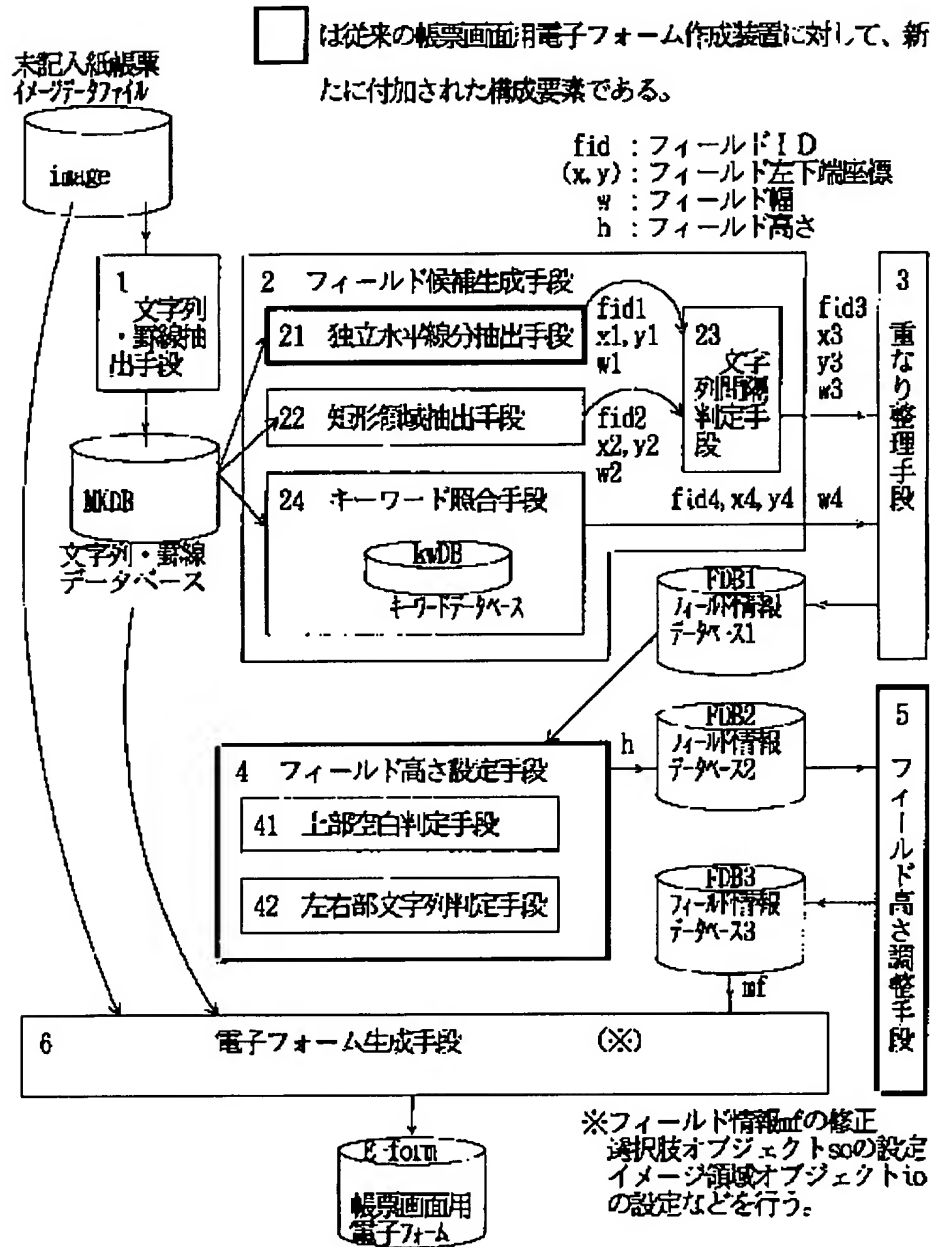


s0 選択肢オブジェクト

io イメージ領域オブジェクト

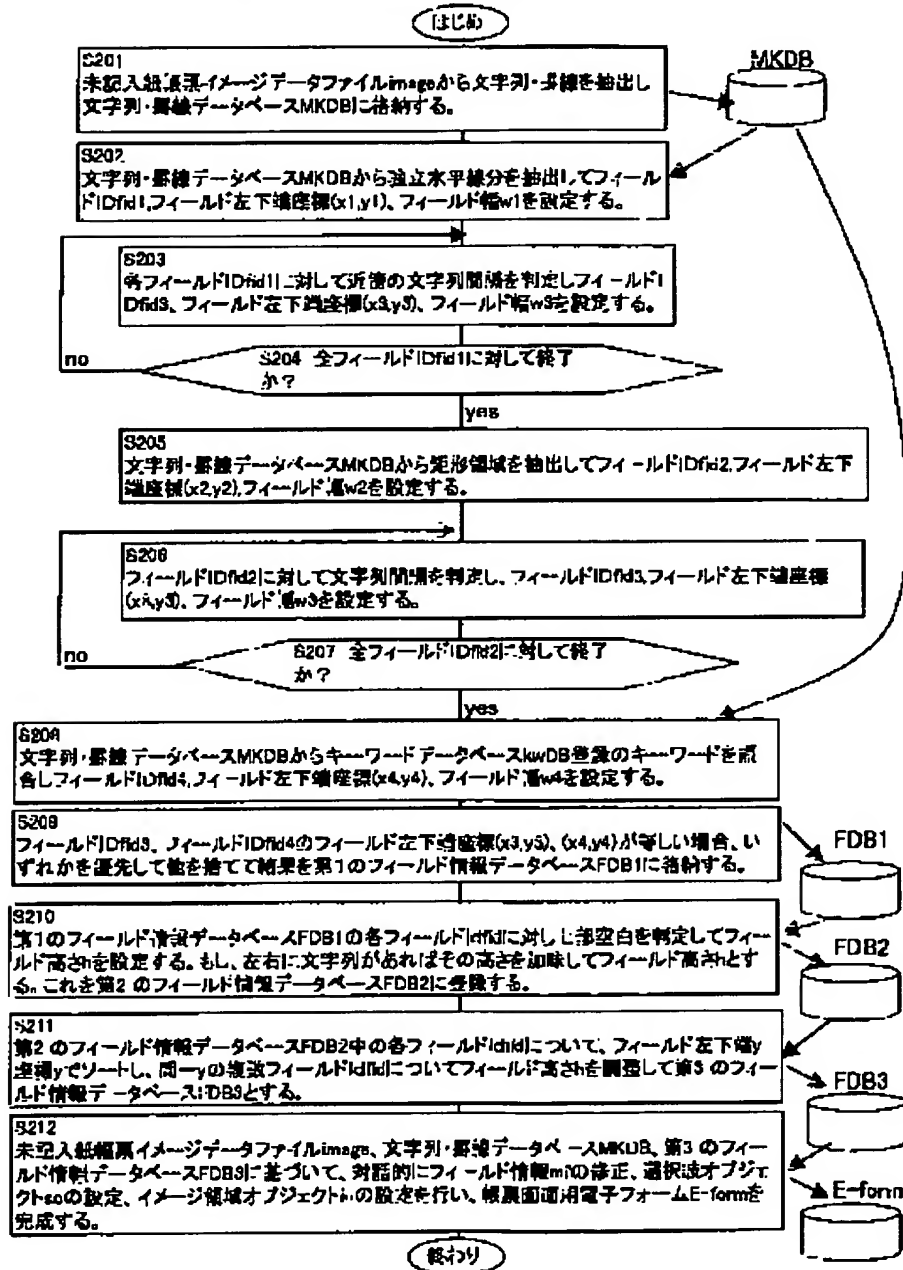
【図1】

## 本発明の帳票画面用電子フォーム作成装置の実施例構成図



【図2】

## 本発明の帳票画面用電子フォーム作成装置の動作フロー図



【図3】

## 文字列・罫線データベースMDBのデータ構造例

(a) 図7のB部の文字列Which?の場合

```

<draw name="FFText241" y="516.47mm" x="71.12mm" w="27.60mm" h="2.12mm">
  <value>
    <text>Which?</text>
  </value>
  .
  .
  .
</font>
</draw>

```

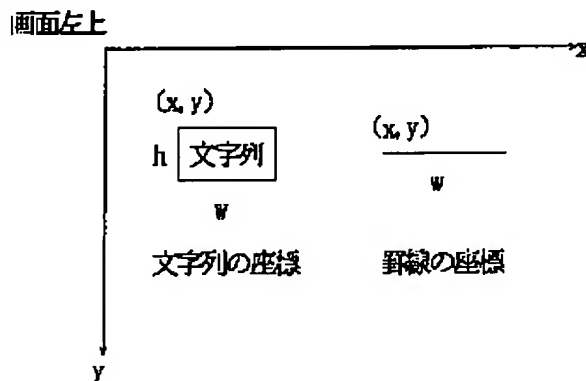
(b) 図7のB部の罫線b1の場合

```

<draw name="FFLine109" y="523.58mm" x="103.80mm" w="40.81mm" h="0.00mm">
  <value>
    <line hand="right" slope="Y">
      .
      .
      .
    </line>
  </value>
</draw>

```

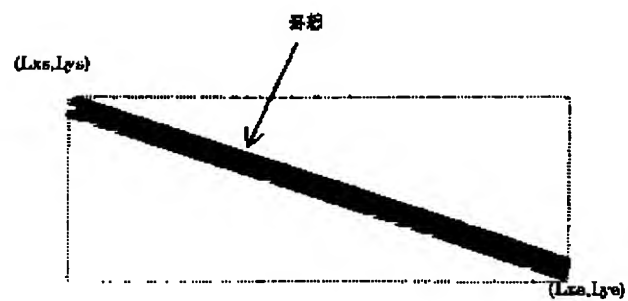
本文字列・罫線データベースMDBの例では座標系を下图のようにとり、文字列の位置座標(x,y)は文字列の左上端の点座標(x,y)でとられている。



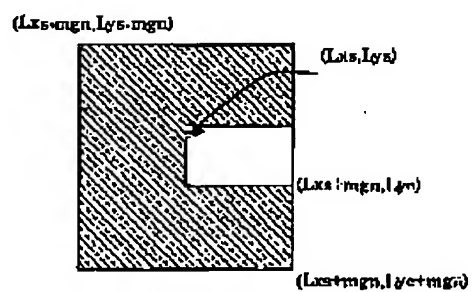
【図1】

独立水平線分抽出手段 21 の動作例

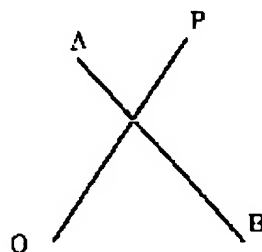
(a) 罫線情報の例



(b) 罫線左端の調査範囲の例



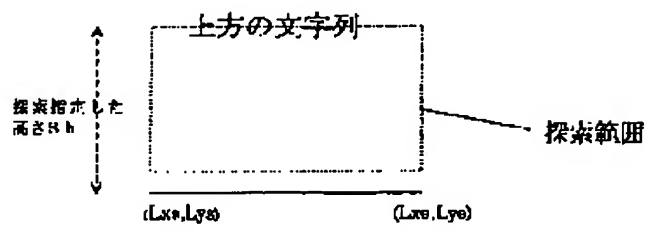
(c) 2線分の交差判定の説明



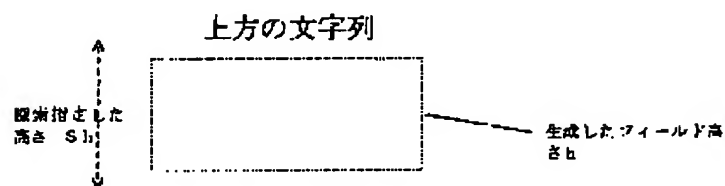
【図5】

上部空白判定手段 41 の動作例

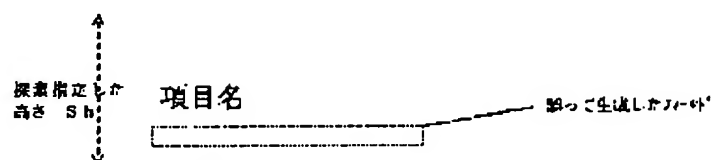
(a) 探索範囲の設定



(b) 生成したフィールド高さ  $h$  の例



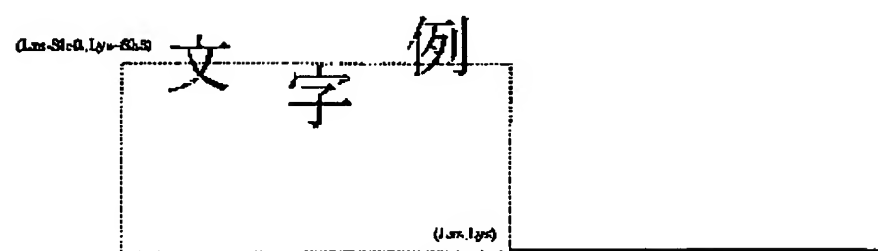
(c) 誤って生成したフィールドの例



【図6】

左右部文字列判定手段 42 の動作例

(a) 左方調査範囲の設定



(b) 生成されたフィールドの例



【岡7】

紙様例



FEE: \$ 20.00

The Commonwealth of Massachusetts  
Executive Office of Public Safety  
State Boxing Commission

FOR ADMINISTRATION USE ONLY!  
DO NOT WRITE IN THIS AREA!

This License was granted:  
Date: \_\_\_\_\_  
Expires: \_\_\_\_\_  
License No.: \_\_\_\_\_

## APPLICATION FOR LICENSE AS A BOXER

IN ACCORDANCE WITH THE PROVISIONS OF CHAPTER 147, GENERAL LAWS, AND RULES AND REGULATIONS OF THE MASSACHUSETTS STATE BOXING COMMISSION

DATE \_\_\_\_\_, 19\_\_\_\_

I hereby make application for a license to act as a professional boxer at boxing, sporting matches or exhibitions

(Please Print With Ball Point Pen)

Name		Assumed or "Ring" Name	
Address		Telephone No. ( )	
City	State	Zip	Country

ASP

DATE OF BIRTH: Mar. \_\_\_\_ Day \_\_\_\_ Yr. \_\_\_\_ PLACE BORN: City \_\_\_\_ State \_\_\_\_ Country \_\_\_\_

HEIGHT: \_\_\_\_ ft. \_\_\_\_ in. WEIGHT: \_\_\_\_ lbs. COLOR EYES: \_\_\_\_ HAIR: \_\_\_\_

COMPLEXION: \_\_\_\_ DISTINGUISHING MARKS: \_\_\_\_

OCCUPATION: \_\_\_\_ EMPLOYER: \_\_\_\_

EMPLOYER ADDRESS: \_\_\_\_ TELEPHONE NO. ( ) \_\_\_\_

CITY \_\_\_\_ STATE \_\_\_\_ ZIP \_\_\_\_ COUNTRY \_\_\_\_

Have you ever held a License to Box in Massachusetts? YES : NO

Have you ever been a licensed Boxer in other states? YES NO

Which? b1 b2 b3 b4 b5 b6 b7 b8 b9 } B&amp;P

Amateur Record: W \_\_\_\_ L \_\_\_\_ Years 19 \_\_\_\_ to 19 \_\_\_\_

Pro Record: W \_\_\_\_ L \_\_\_\_ Years 19 \_\_\_\_ to 19 \_\_\_\_

Have you ever been convicted of a felony in the past ten (10) years? YES NO IF YES, please provide details:

Date \_\_\_\_ Offense \_\_\_\_ Court \_\_\_\_ C 3 C 4 } C&amp;P

Have you ever been convicted of a misdemeanor in the past five (5) years? YES NO IF YES, please provide details:

Date \_\_\_\_ Offense \_\_\_\_ Court \_\_\_\_ C 3 C 4

A TRUE STATEMENT MADE UNDER THE PENALTIES OF PERJURY

\* Signature of Applicant \_\_\_\_\_



【図8】

紙帳票と帳票画面の対比例

(a) 紙帳票例

年月日

XXX申請書

住所

都府道県 市郡 区

a81

(b) 帳票画面例

年月日

XXX申請書

住所

都府道県 市 区

b81 b82 b83 b84

b85 b86 b87 b88

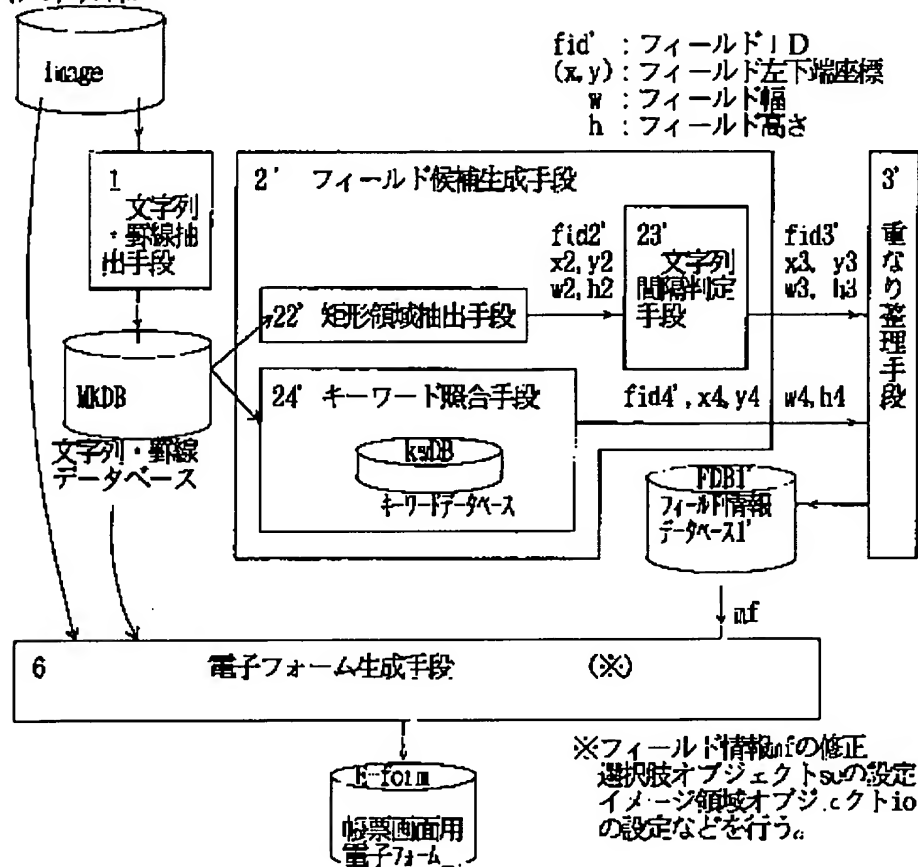
SO

スカーン選択肢打テ外

【図9】

従来の帳票画面用電子フォーム作成装置構成例とフィールド情報のデータ構造例

(a)従来の帳票画面用電子フォーム作成装置構成例  
未記入紙帳票  
イメージファイル



(b)フィールド情報nfのデータ構造例

フィールドID	フィールド左下端座標	フィールド幅	フィールド高さ	フィールド名称	属性	..
fid	(x,y)	w	h	xxxxxxx	X(n)	..
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.

フロントページの続き

F-Form (参考) 5L096 A407 B417 E427 F403 F410  
F413 F416 F454 F464 F466  
F467 F469